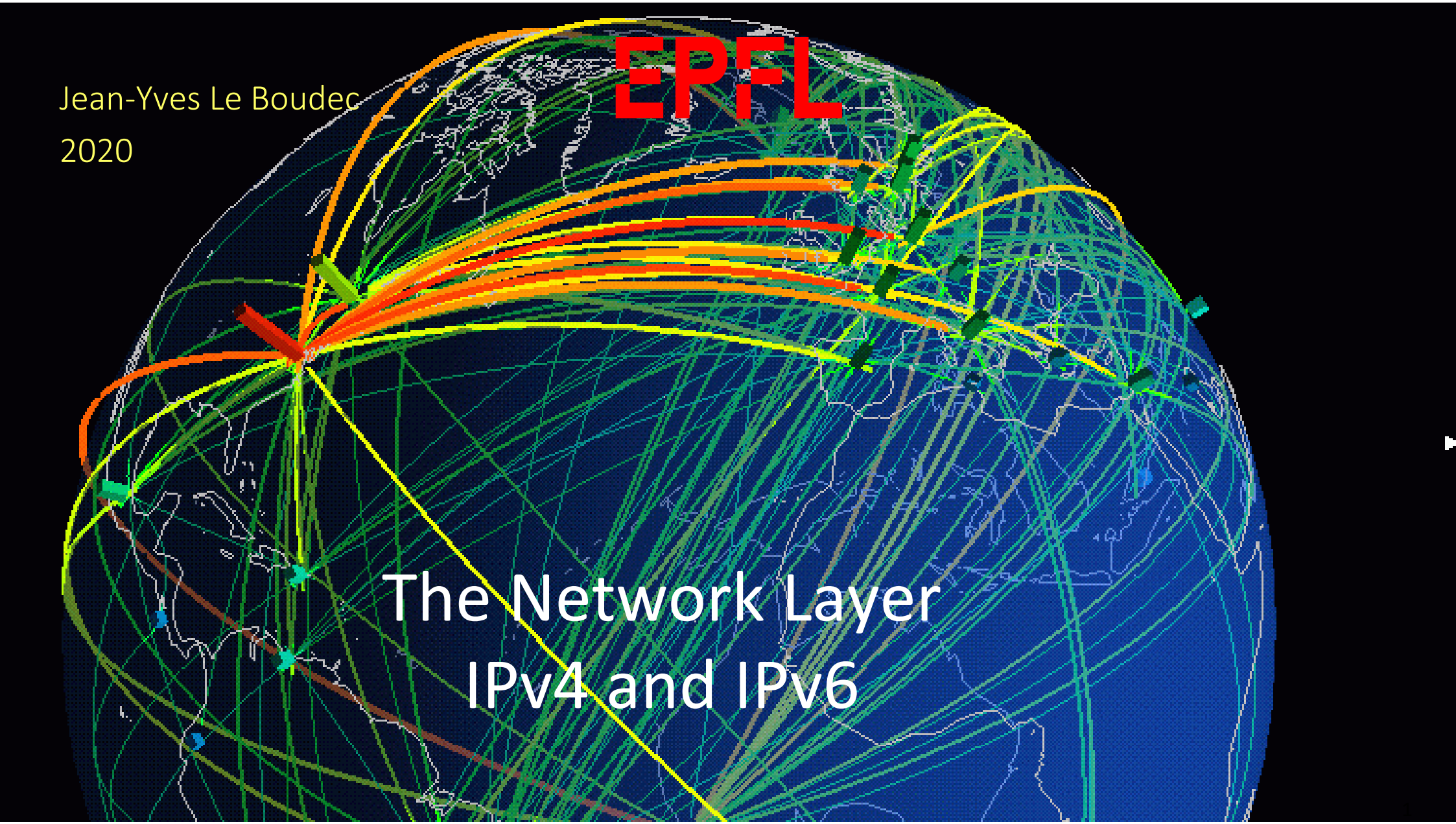
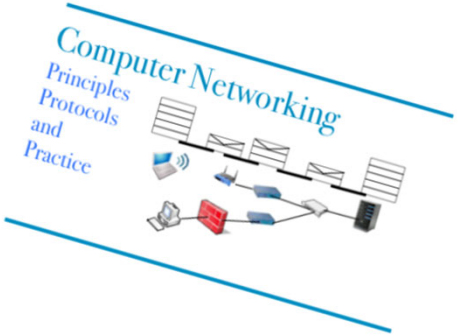

EPFL
Jean-Yves Le Boudec
2020
The Network Layer
IPv4 and IPv6

Contents

1. The Two Principles of IP Unicast
2. IPv4 addresses
3. IPv6 addresses
4. NATs
5. ARP
6. Host configuration
7. Hop Limit and TTL

Textbook

Chapter 5: The Network Layer

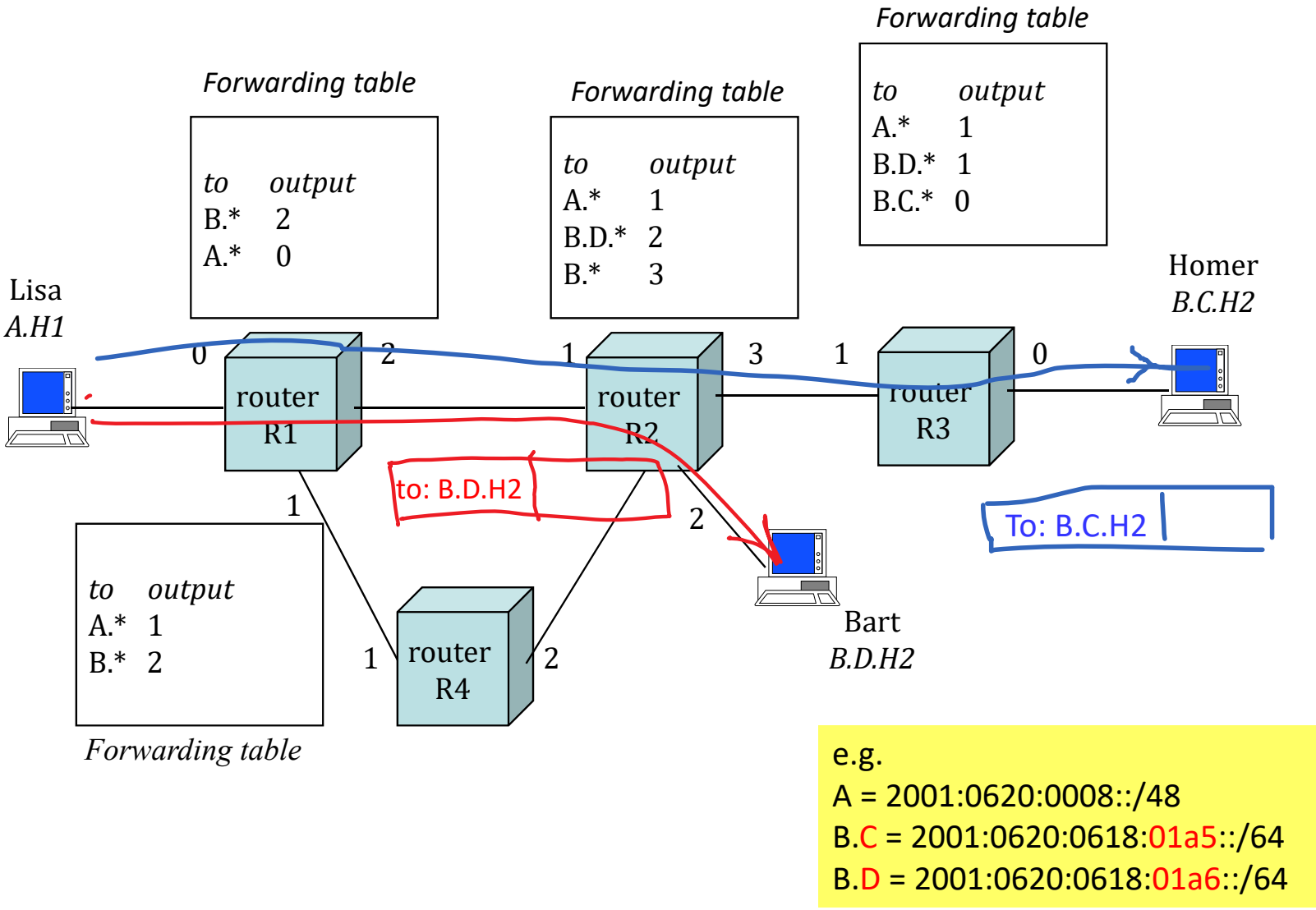


IP Principle #1 = Structured addresses + Longest prefix match

Recall goal of Internet Protocol (IP) = interconnect all systems in the world

Principle #1:

- every interface has an IP address
- IP address reflects which network the system belongs to
- every packet contains IP address of destination
- every system (host = end-system, router = intermediate system) has a forwarding table (= routing table) and performs longest prefix match on destination address



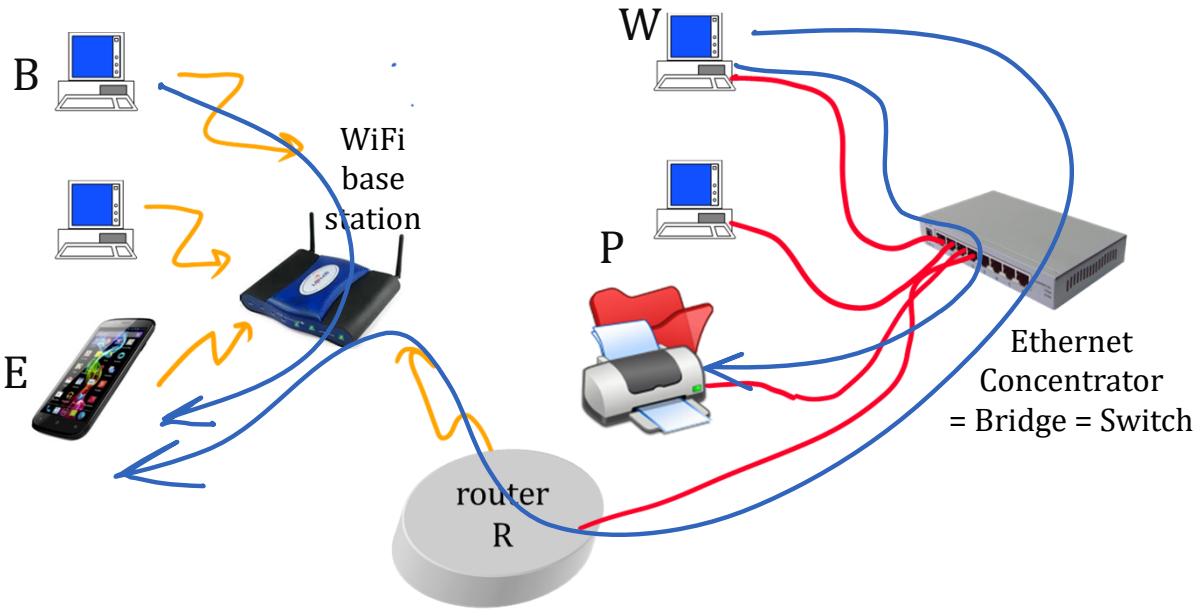
e.g.
A = 2001:0620:0008::/48
B.C = 2001:0620:0618:01a5::/64
B.D = 2001:0620:0618:01a6::/64

IP Principle #2 =
Don’t use routers inside a LAN

B ↔ E and W↔P should not go through router

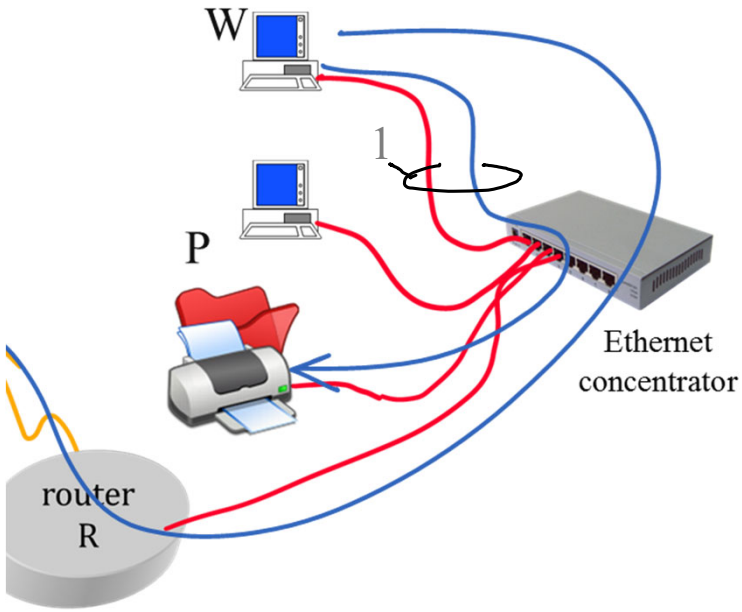
W ↔ E goes through router

IP principle #2 says:
between subnets (= LANs)
use routers, inside subnet don’t



Hosts in same subnet must have same subnet mask and same subnet prefix

We observe a packet from W to P at 1. Which IP destination address do we see ?



A. The IP address of P
B. The IP address of an Ethernet interface of the Ethernet concentrator
C. There is no destination IP address in the packet since communication is inside the subnet and does not go through a router
D. I don't know

Solution

Answer A

The IP header is always present if we use TCP/IP, even if communication is inside the same LAN.

2. IPv4 addresses

IPv4 address: 32 bits, usually noted in dotted decimal notation
Uniquely identifies one interface in the world (in principle)

dotted decimal: 4 integers (one integer = 8 bits)

example 1: `128.191.151.1`

example 2: `129.192.200.2`

hexadecimal: 8 hex digits (one hex digit = 4 bits)

example 1: `x80 bf 97 01`

example 2: `x81 c0 c8 02`

binary: 32 bits

example 1: `b1000 0000 1011 1111 1001 0111 0000 0001`

example 2: `b1000 0001 1100 0000 1100 1000 0000 0010`

Binary, Decimal and Hexadecimal

Given an integer B (the basis) any integer can be represented as a string in an alphabet of B symbols

	Basis	*Alphabet*	*Example*
Decimal	X	$\{0,1,2,3,4,5,6,7,8,9\}$	200
Binary	II	$\{0,1\}$	$1100\ 1000$
Hexadecimal	XVI	$\{0,1,2,3,4,5,6,7,8,9,a,b,c,d,e,f\}$	$c8$

Binary <-> hex is easy: one hex digit (= nibble) is 4 binary digits

$c_{hex} = 1100_{bin}$ $8_{hex} = 1000_{bin}$ $c8_{hex} = 1100\ 1000_{bin}$

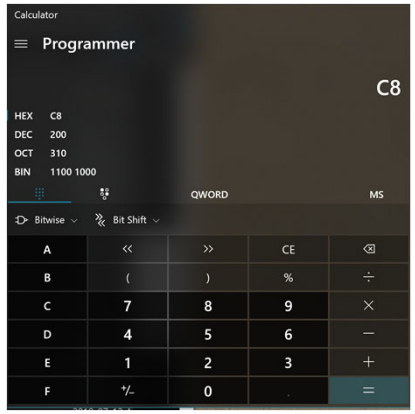
Binary/ hex <-> decimal is best done by a calculator

$1100\ 1000_{bin}$ = 128 + 64 + 8 = 200

Special Cases to remember

$f_{hex} = 1111_{bin} = 15$

$ff_{hex} = 1111\ 1111_{bin} = 255$



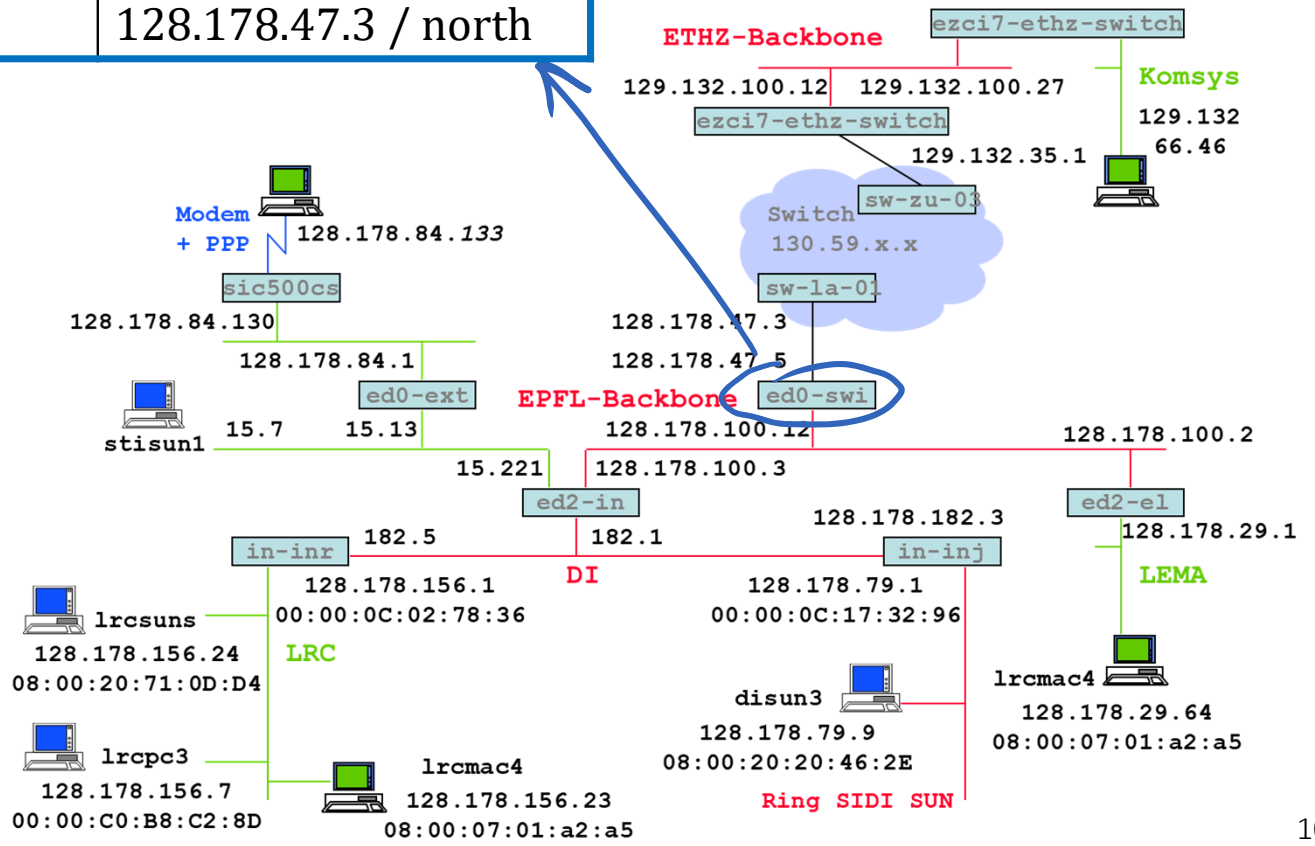
Example: Routing Table at ed0-swi

Destination	*Next-Hop / Interface*
128.178.29/24	128.178.100.2 / south
128.178/16	128.178.100.3 / south
0/0	128.178.47.3 / north

/24 is prefix length in bits

0/0 (empty string) means default route

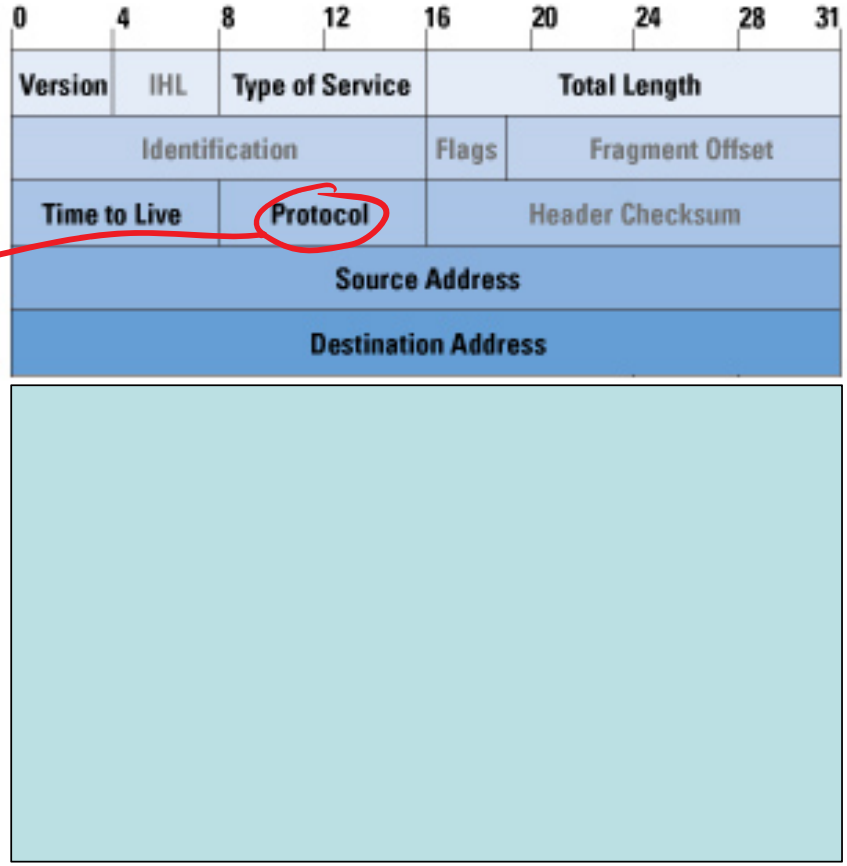
to 128.178.* → to ed2-in
to 128.178.29.* → to ed2- el
else → north



Special Addresses

0.0.0.0	absence of address
127.0.0/24 for example 127.0.0.1	this host (loopback address)
10.0.0.0/8, 172.16.0.0/12, 192.168.0.0/16	private networks (e.g in IEW) cannot be used on the public Internet
169.254.0.0/16	link local address (can be used only between systems on same LAN)
224/4	multicast
240/5	reserved
255.255.255.255/32	link local broadcast

IPv4 Packet Format



Higher layer protocol 1= ICMP, 6 = TCP, 17 = UDP)

Header 20 bytes (+ options, if any)

payload

The mask 255.255.254.0 means that the subnet is made of the first …

A. 16 bits
B. 18 bits
C. 22 bits
D. 23 bits
E. 24 bits
F. I don’t now

Solution

Answer D

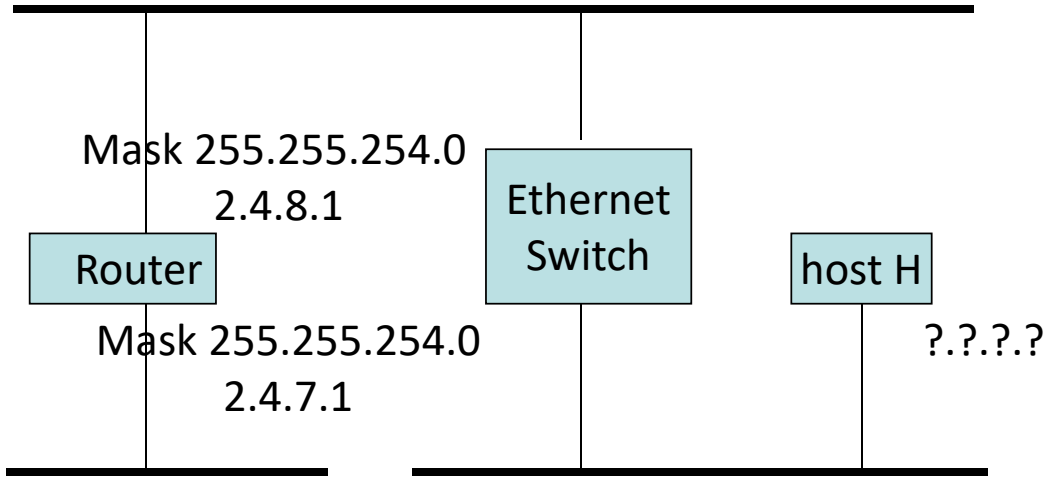
254 = 0b 1111 1110 i.e.

255.255.254.0 = 0b 1111 1111 1111 1111 1111 1110 0000 0000

23 bits equal to 1

Which address is a valid choice for H ?

A. 2.4.8.2
B. 2.4.9.1
C. Both A and B
D. None
E. I don't know



Solution

Answer C

Router north and H are in the same subnet so H has a subnet prefix of 23 bits.

Both answers A and B have same subnet prefix as router north:

2.4.8 = 0000 0010 0000 0100 0000 1000
Router north's subnet prefix: 2.4.8 / 23 = 0000 0010 0000 0100 0000 100

A's subnet prefix is 2.4.8/23

B's subnet prefix is 2.4.9/23

2.4.9 = 0000 0010 0000 0100 0000 1001

2.4.9 / 23 = 0000 0010 0000 0100 0000 100

2.4.9/23 = 2.4.8/23 !

3. IPv6 Addresses

The old version of IP is IPv4. IPv6 is the current (probably final) version of IP

Why a new version ?

IPv4 address space is too small (32 bits $\rightarrow \approx 4 \cdot 10^9$ addresses)

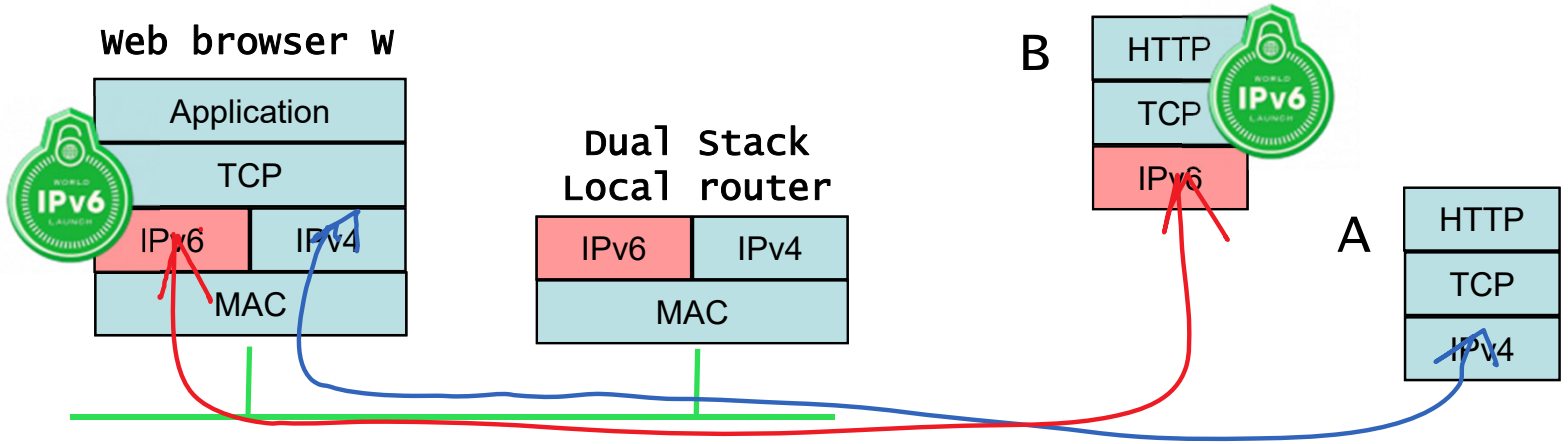
What does IPv6 do ?

Redefine packet format with a larger address: 128 bits ($\approx 3 \cdot 10^{38}$ addresses)

Otherwise essentially the same as IPv4

But IPv6 is incompatible with IPv4; routers and hosts must handle both separately

A can talk to W, B can talk to W, A and B cannot communicate at the network layer



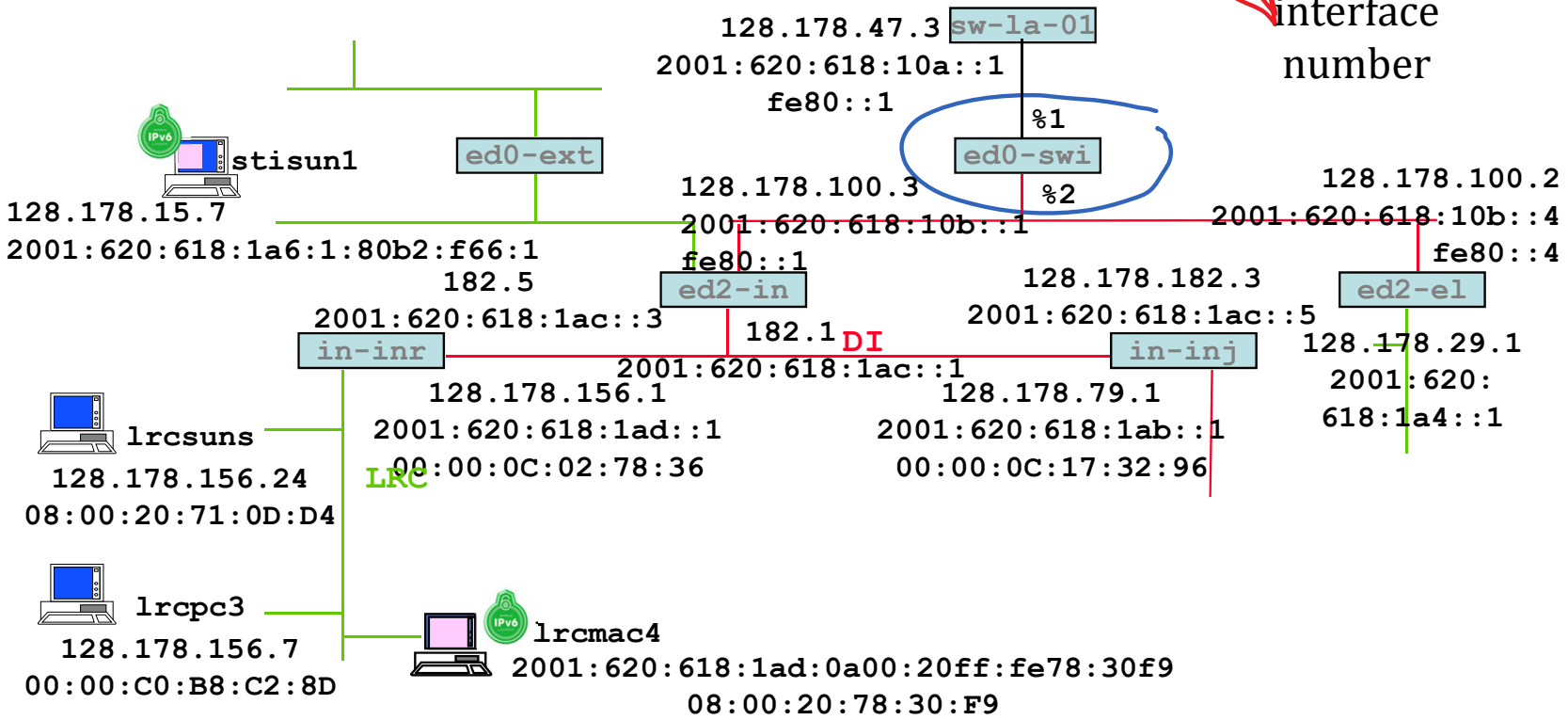
Routing Tables at ed0-swi

Destination	Next hop
2001:620:618:1a4/64	fe80::1%2
2001:620:618/48	fe80::4%2
::/0	fe80::1%1

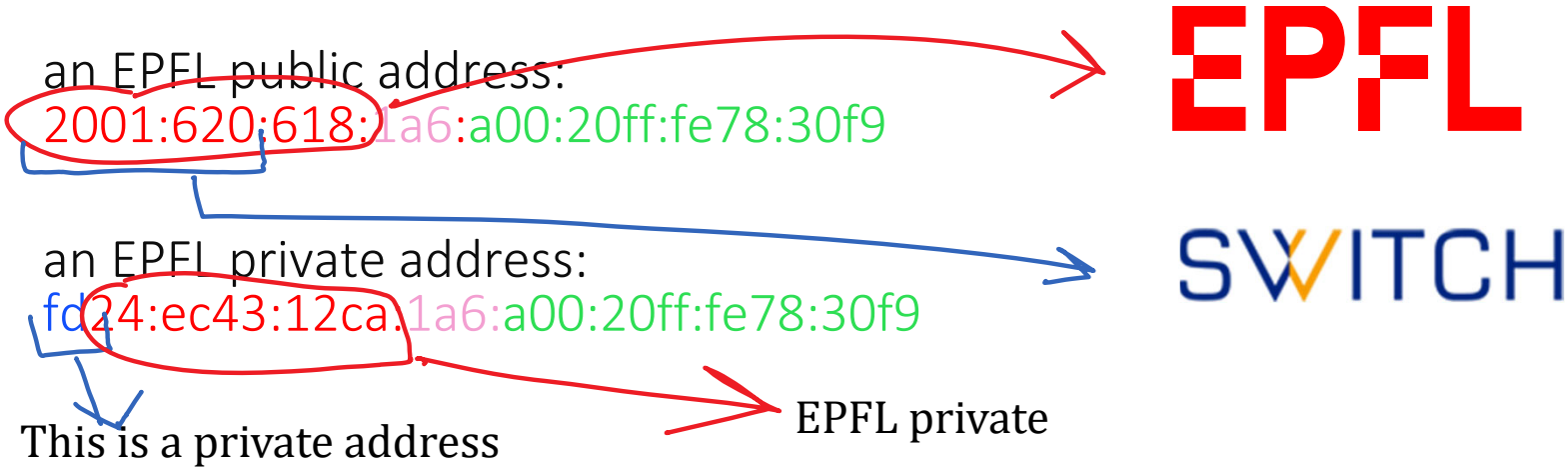
Destination	Next hop
128.178.29/24	128.178.100.2
128.178/16	128.178.100.3
0/0	128.178.47.3

IP address of next hop

interface number



IPv6 addresses are 128 bit long and are written using hexadecimal digits



Compression Rules for IPv6 Addresses

1 *hextet* = 1 *piece* = 16 bits = [0-4]hex digits;

prefer lower case

pieces separated by “:” (colon)

one IPv6 address uncompressed = 8 pieces

- :: replaces any number of 0s in more than one piece; appears at most once in address
- leading zeroes in one piece are omitted ;

uncompressed	*compressed*
2002:0000:0000:0000:0000:ffff:80b2:0c26	2002::ffff:80b2:c26
2001:0620:0618:01a6:0000:20ff:fe78:30f9	2001:620:618:1a6:0:20ff:fe78:30f9

A Few IPv6 Global Unicast Addresses

The block 2000/3 (i.e. 2xxx and 3xxx) is allocated for global unicast addresses

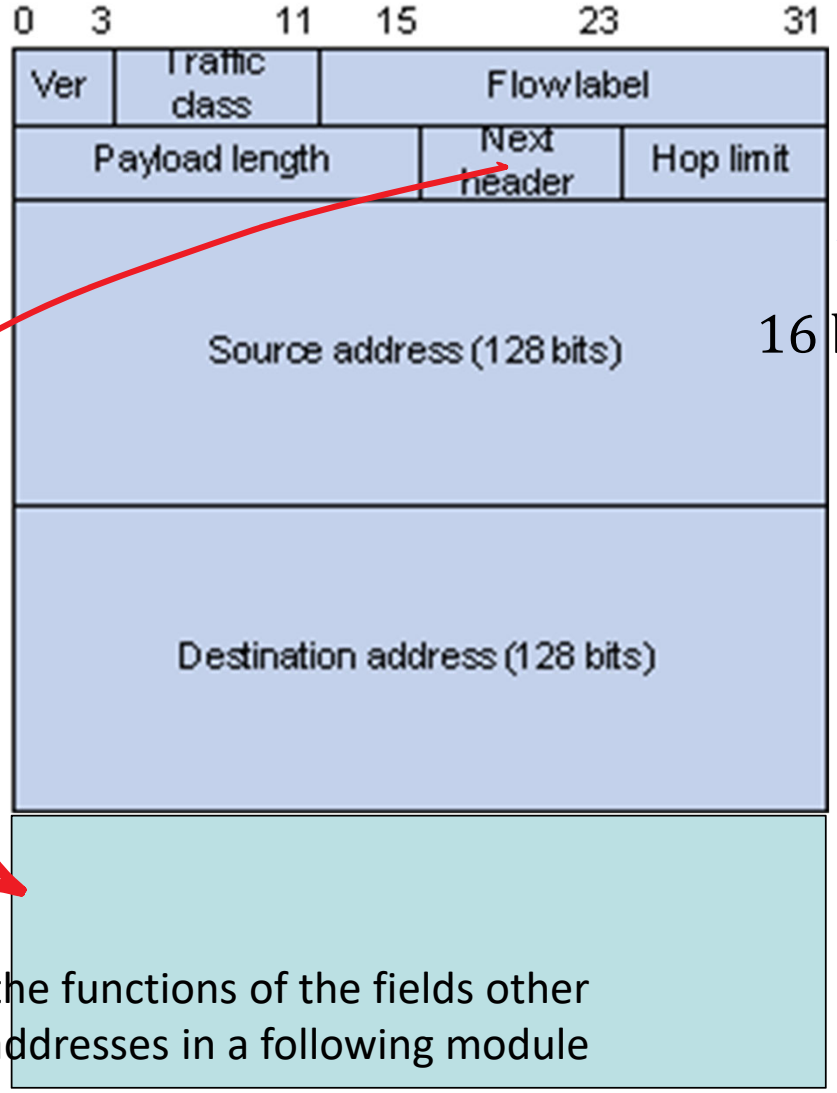
2001:620::/32	Switch
2001:620:618::/48	EPFL
2001:620:8::/48	ETHZ
2a02:1200::/27	Swisscom
2001:678::/29	provider independent address
2001::/32	Teredo (tunnels IPv6 in IPv4)
2002::/16	6to4 (tunnels IPv6 in IPv4)

Examples of Special Addresses

::/128	absence of address
::1/128	this host (loopback address)
fc00::/7 (i.e. fcxx: and fdxx:) For example fd24:ec43:12ca:1a6:a00: 20ff:fe78:30f9	Unique local addresses = private networks (e.g in IEW) cannot be used on the public Internet
fe80::/10	link local address (can be used only between systems on same LAN)
ff00::/8	multicast
ff02::1:ff00:0/104	Solicited node multicast
ff02::1/128 ff02::2/128	link local broadcast all link local routers

EPFL Private

IPv6 Packet Format



e.g. Higher layer protocol 1= ICMP, 6 = TCP, 17 = UDP)

16 bytes

Header 40 bytes (+ options, if any)

payload

We will see the functions of the fields other than the addresses in a following module

The dotted decimal notation for $0102{:}\,ffff$ is …

A. 1.2.255.255
B. 16.32.255.255
C. 228.393.255.255
D. I don’t know

In full, the hexadecimal notation «2001::bad:babe» means…

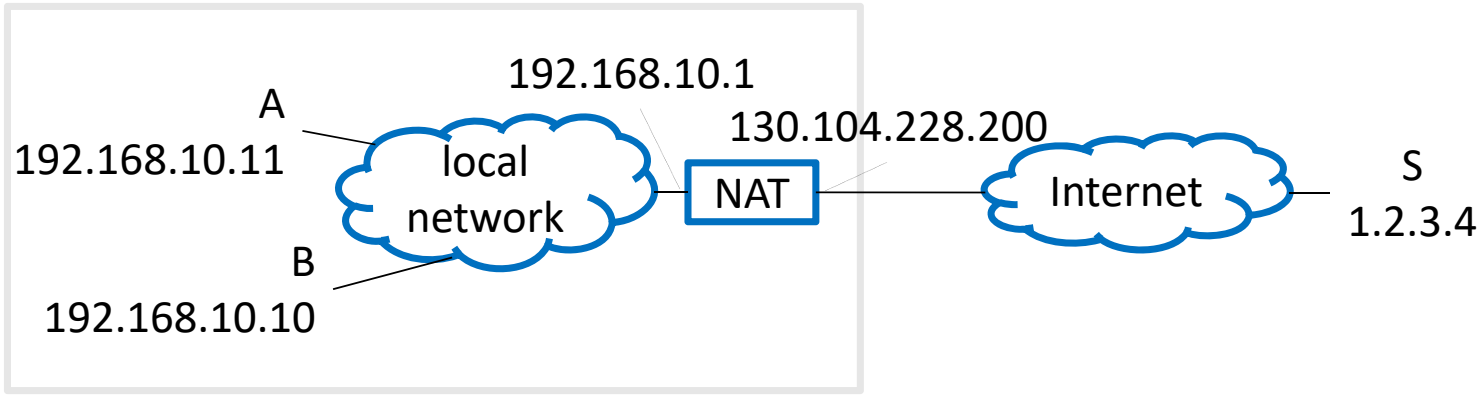
A. 2001:0bad:babe
B. 2001:0000:0000:0000:0000:0000:0bad:babe
C. 2001:0000:0bad:babe
D. 2001:0000:bad:babe
E. None of the above
F. I don’t know

Solution

Answer A
Recall the mapping (hex) ff → (decimal) 255

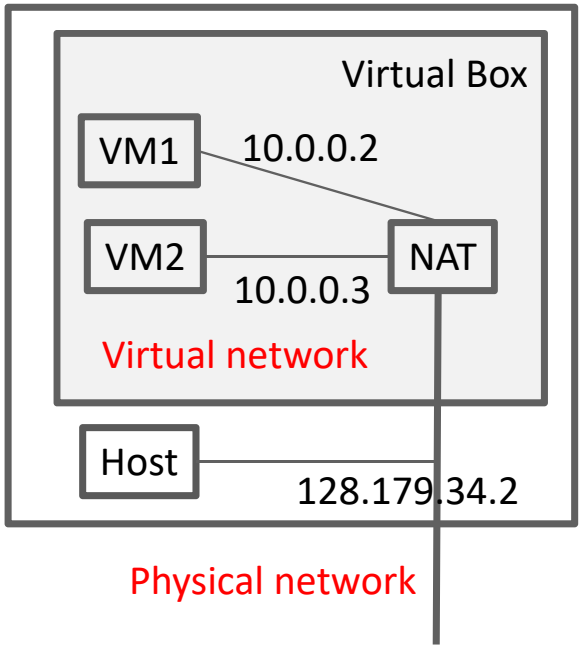
Answer B. The convention :: means as many 0's as required to make the string 128 bits.
Leading zeros are omitted, so that :bad: means :0bad:

4. NAT (Network Address Translation) box
Why invented ?



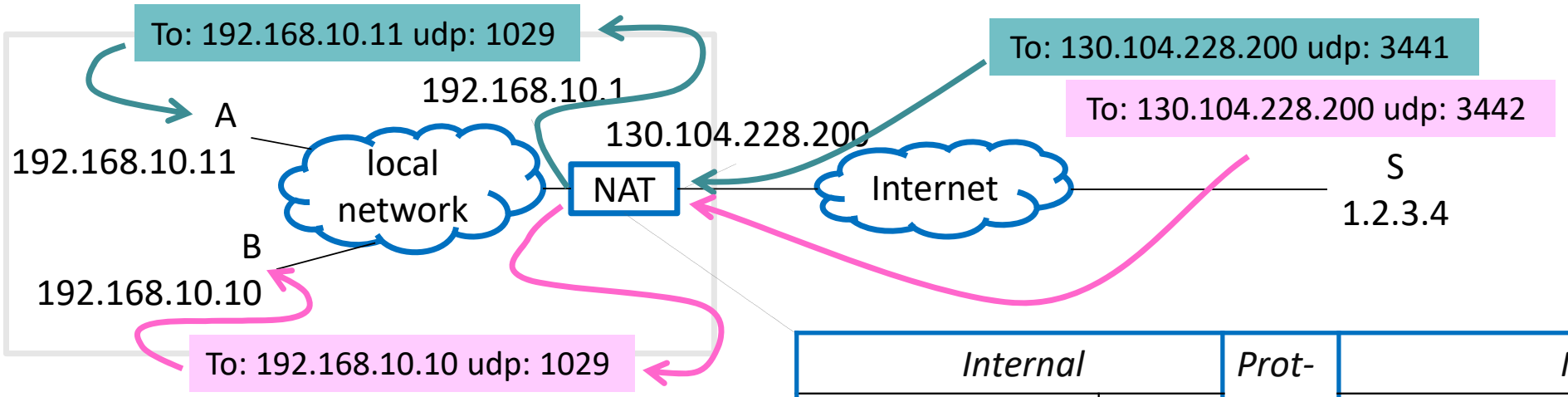
Internet service provider gives you one IP address e.g. but you have n devices at home and need more than 1 IP address.

Virtual box has n guests + one host and wants them to communicate with outside while using one IP address of the physical machine.



Goal of NAT: allow $n > 1$ devices to use one single IP address.
It is our first example of a middle box that deviates from the TCP/IP architecture: it violates: IP principle #1 (+ layering, see next slide).

NAT Translates Internal Addresses and Ports (1)



NAT Table

Internal		*Prot-ocol*	*NAT*	
IP address	*Port*		*IP address*	*Port*
192.168.10.11	1029	udp	130.104.228.200	3441
192.168.10.10	1029	udp	130.104.228.200	3442

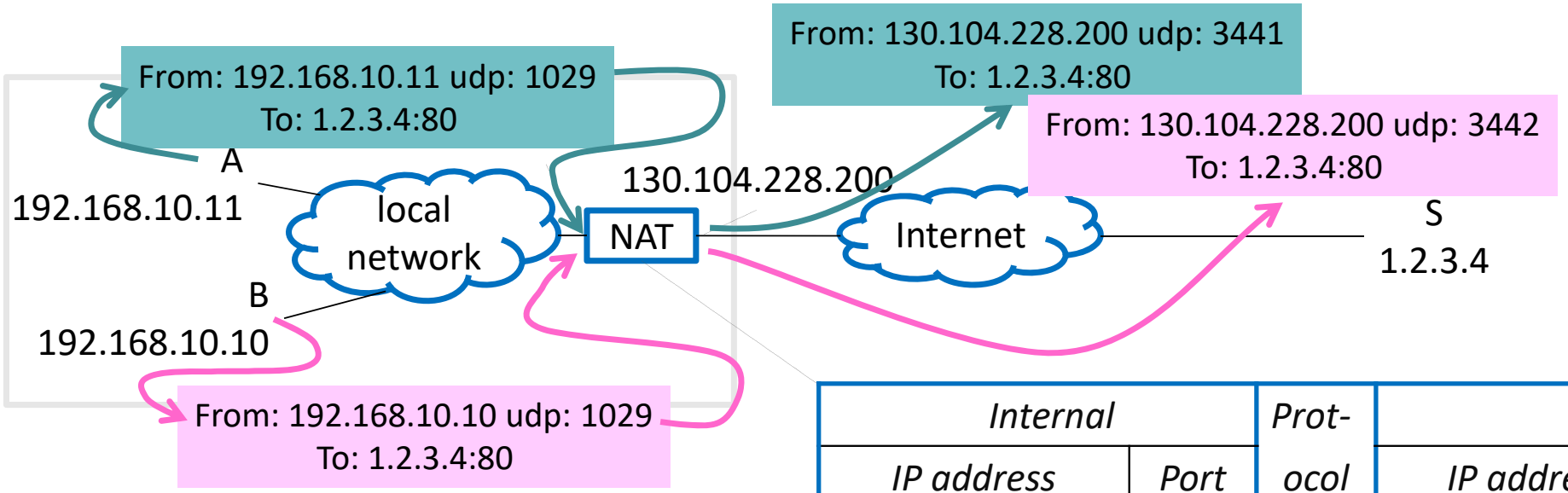
NAT translates (= masquerades)
internal IP address *and* internal port number into NAT IP address and NAT port number

Outside sees only NAT IP address and NAT port

NAT forwarding is based on exact matching in NAT table

Implemented by iptables in Linux

NAT Translates Internal Addresses and Ports (2)



NAT Table

Internal		*Prot-ocol*	*NAT*	
IP address	*Port*		*IP address*	*Port*
192.168.10.11	1029	udp	130.104.228.200	3441
192.168.10.10	1029	udp	130.104.228.200	3442

In packets from external (a.k.a.WAN) to internal (a.k.a. LAN), NAT translates destination address + port.

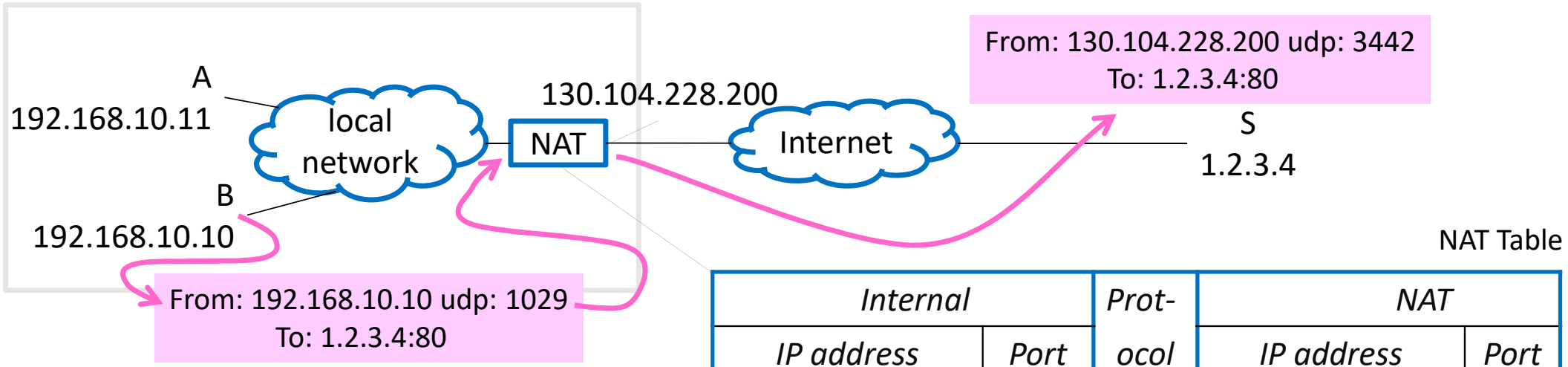
In packets from internal to external, NAT translates source address + port.

Internal addresses are typically private addresses.

How does NAT maintain NAT table ?

NAT creates a NAT table entry on-the-fly (automatically) when client on internal network contacts server on external network.

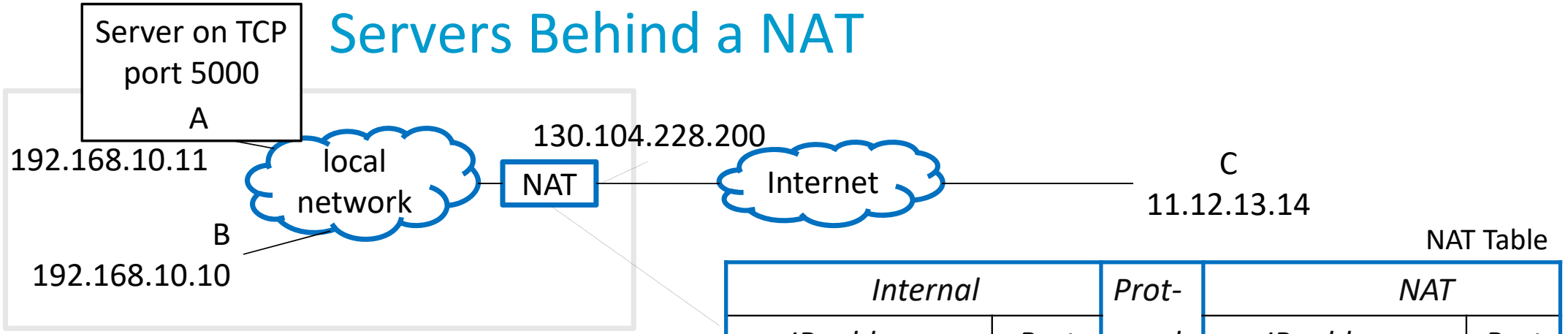
NAT chooses a NAT port that does not create collision in the table.



NAT Table

Internal		*Prot-*	*NAT*	
IP address	*Port*	*ocol*	*IP address*	*Port*
192.168.10.11	1029	udp	130.104.228.200	3441
192.168.10.10	1029	udp	130.104.228.200	3442

Servers Behind a NAT



NAT Table

Internal		*Prot-ocol*	*NAT*	
IP address	*Port*		*IP address*	*Port*
192.168.10.11	1029	udp	130.104.228.200	3441
192.168.10.10	1029	udp	130.104.228.200	3442
192.168.10.11	5000	tcp	130.104.228.200	5000

Assume A has a server on tcp port 5000

Problem: Automatic operation of NAT requires communication to be started by A, which is not done for a server

Solution: manual configuration of port forwarding in NAT

C connects to A at 130.104.228.200 port 5000.

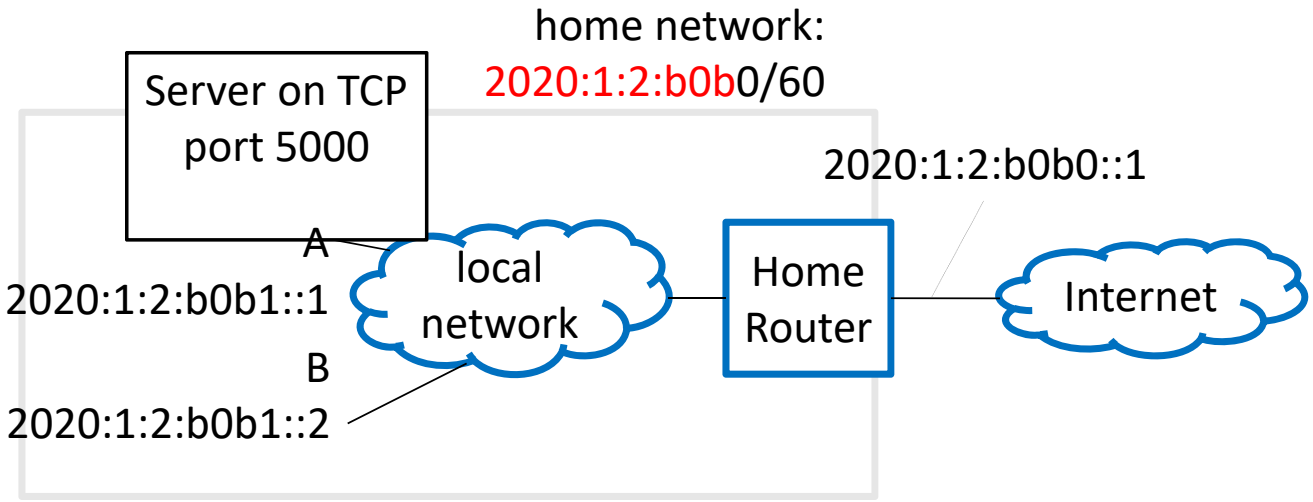
A still needs to know its NAT IP address and advertise it to potential clients like C. A discovers its NAT IP address if A and NAT use e.g. UPnP.

Setup by manual configuration of *port forwarding* on NAT

This provides protection to home network: a server port can be accessed only if explicitly configured. Servers on the internet are exposed to infections and need to be actively protected.

NATs and IPv6

NAT was developed for IPv4, motived by lack of IPv4 addresses; it also exists for IPv6 but is little used.

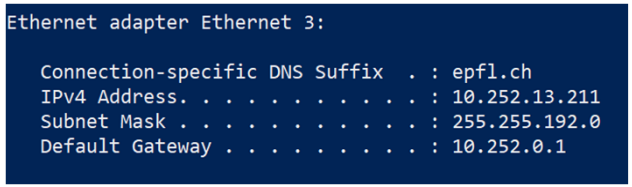


With IPv6, home routers often do not use NAT because provider typically allocates a block of IPv6 addresses, not just one as with IPv4.

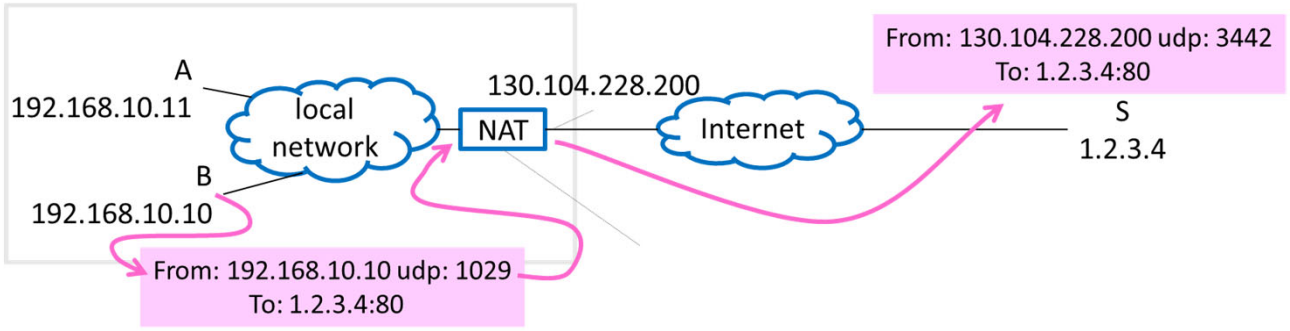
With IPv6 the home router provides protection by acting as a filtering router: allows communication from outside only if initiated from inside, unless manually configured.

More on NATs

- In the previous slides, the NAT mapping
  (internal addr, internal port) → (NAT addr, NAT port)
  is independent of the external correspondent (such as S).
  Such a NAT («full cone») is friendly to external correspondents. The mapping can be learnt by S (e.g. a skype server) and used by any other external correspondent S' (e.g. a skype user). Good for peer-to-peer communication: after setup, skype calls use direct communication, without server – no store-and-forward.
  Not all NATs are full cone. E.g. with a «symmetric» NAT, the NAT mapping is
  (internal addr and port, external addr and port) → (NAT addr, NAT port)
  The mapping learnt by one external correspondent S cannot be used or guessed by another S'. Skype must use a store-and-forward server during entire call.
- NATs can be concatenated: Virtual Box NAT in home NAT
- "Carrier-grade NATs" share p public addresses with $n > p$ internal hosts ("carrier-grade NAT").
  Example: EPFL VPN uses 10/8 addresses and a NAT.

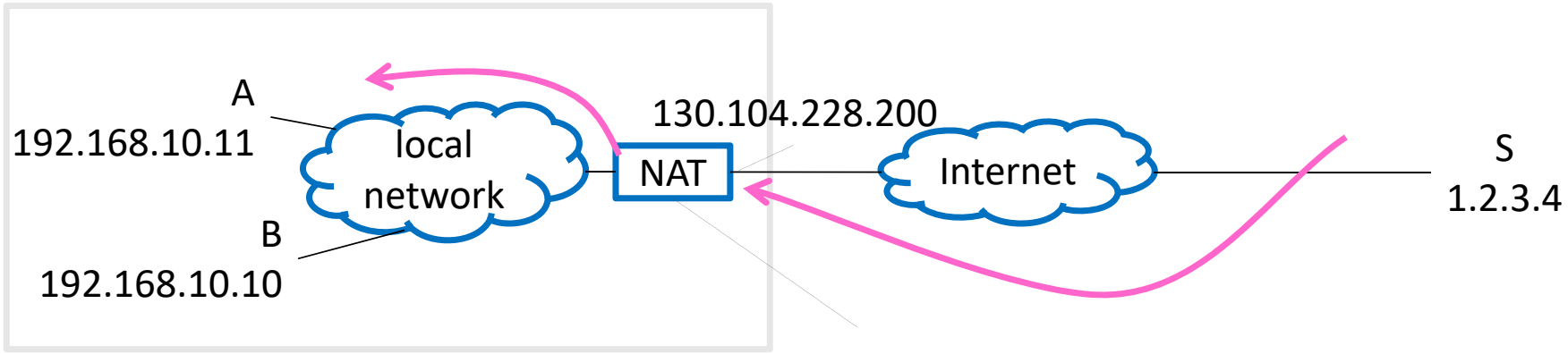


When a NAT has a packet to forward and an association exists in the NAT table…



A. The NAT looks for a longest prefix match
B. The NAT looks for an exact match
C. None of the above
D. I don’t know

From WAN to LAN, the NAT may modify…



A. The source port
B. The destination port
C. None of the above
D. I don’t know

Solution

Answer B in both cases.

5. MAC Address Resolution

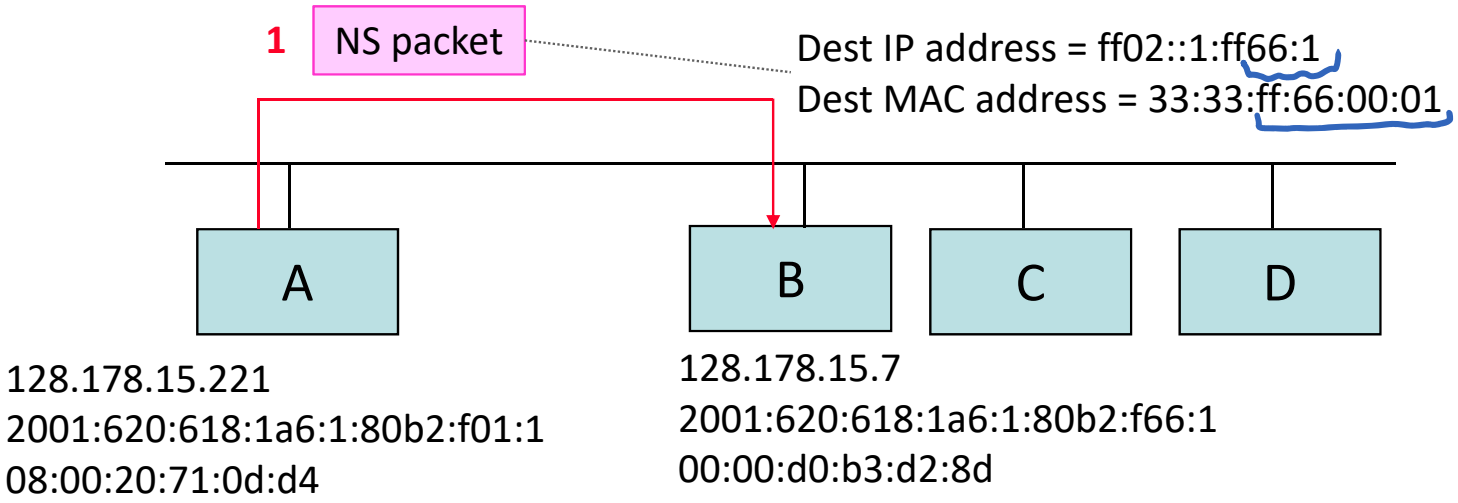
Why ?

An IP machine A has a packet to send to a next-hop B (final destination or next-hop router). A knows B's IP address; A must find B's MAC address.

How ?

On Ethernet, A sends an address resolution packet on the LAN. All hosts that have the IP address of B (in principle only B) respond with their MAC address.

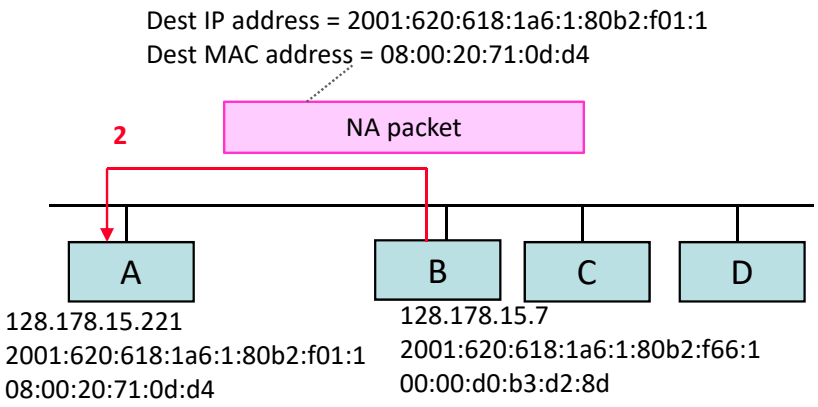
MAC Address Resolution with IPv6 (NDP)



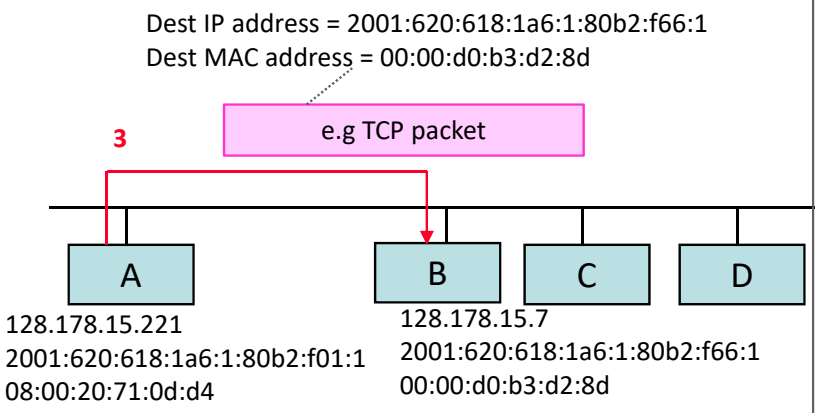
A has a packet to send to B = 2001:620:618:1a6:1:80b2:f66:1

This address is on the same subnet therefore A sends directly to B and looks for B's MAC address

1. A sends a Neighbour Solicitation (NS) packet using the Neighbour Discovery Protocol (NDP) containing the question: "who has IP address B ?". The IP destination address of this packet is a special multicast address (Solicited Node Multicast Address). The MAC address is derived from the multicast IP address. The NS packet is received by all hosts whose IP address has the same 24 bits as B (see later slide).



2. B responds with a Neighbour Advertisement (NA) packet, giving its MAC address. This NA packet is sent by B to A.



3. A reads NA, stores MAC address in its neighbour cache (also called ARP table) and can send the data to B. The cache is refreshed whenever A receives a packet from B; it expires after e.g. 20 mn of inactivity.

NA, NS packets are carried as ICMPv6 packets, next-header =58 (0x3a), inside IPv6 packets.

The Solicited Node Multicast Address

NDP and other protocols use this multicast address - obtained by adding last 24 bits of target IP address to ff02::1:ff00:0/104

A packet with such a destination address is forwarded by layer 2 to all nodes that listen to this multicast address

Only for IPv6 – IPv4 uses broadcast instead

Target address	Compressed	2001:620:618:1a6:001:80b2:f66:1
	Uncompressed	2001:0620:0618:01a6:0001:80b2:0f66:0001
Solicited Node multicast address	Uncompressed	ff02:0000:0000:0000:0000:0001:ff66:0001
	Compressed	ff02::1:ff66:1

Look Inside an NDP Neighbour Solicitation Packet

```
ETHER:  Packet size =  86 bytes
ETHER:  Destination = 33:33:ff:01:00:01
ETHER:  Source = 3c:07:54:3e:ab:f2
ETHER:  Ethertype = 0x86dd
ETHER:
IP:   ----- IP Header -----
IP:
IP:   Version = 6
IP:   Traffic class =0x00000000
IP:      ....  0000 00.. .... .... .... .... .... = Default Differentiated Service Field
IP:      ....    ....     ..0.  .... .... .... .... .... = No ECN-Capable Transport (ECT)
IP:      ....    ....     ...0  .... .... .... .... ....  = No ECN-CE
IP:   .... .... ....  0000 0000 0000  0000 0000 =  Flowlabel: 0x00000000
IP:   Payload length = 32
IP:   NextHeader= 58
IP:   Hop limit= 255
IP:   Source address = 2001:620:618:197:1:80b2:97c0:1
IP:   Destination address = ff02::1:ff01:1
IP:
ICMPv6:    ----- ICMPv6 Header -----
ICMPv6:
ICMPv6:  Type = 135
ICMPv6:  Code=0
ICMPv6:  Checksum = 0xb199 [correct]
ICMPv6:  Reserved = 00000000
ICMPv6:  Target Address=2001:620:618:197:1:80b2:9701:1
ICMPv6:
```

Solicited Node Multicast Address corresponding to this IPv6 target address

Neighbor Solicitation (=ARP Request)

MAC Address Resolution with IPv4

Similar, except

the protocol is called ARP (Address Resolution Protocol)
ARP packets are not IP packets but directly in Ethernet frame with Ethertype = ARP (86dd)

NDP NS /NA are called ARP Request /ARP reply

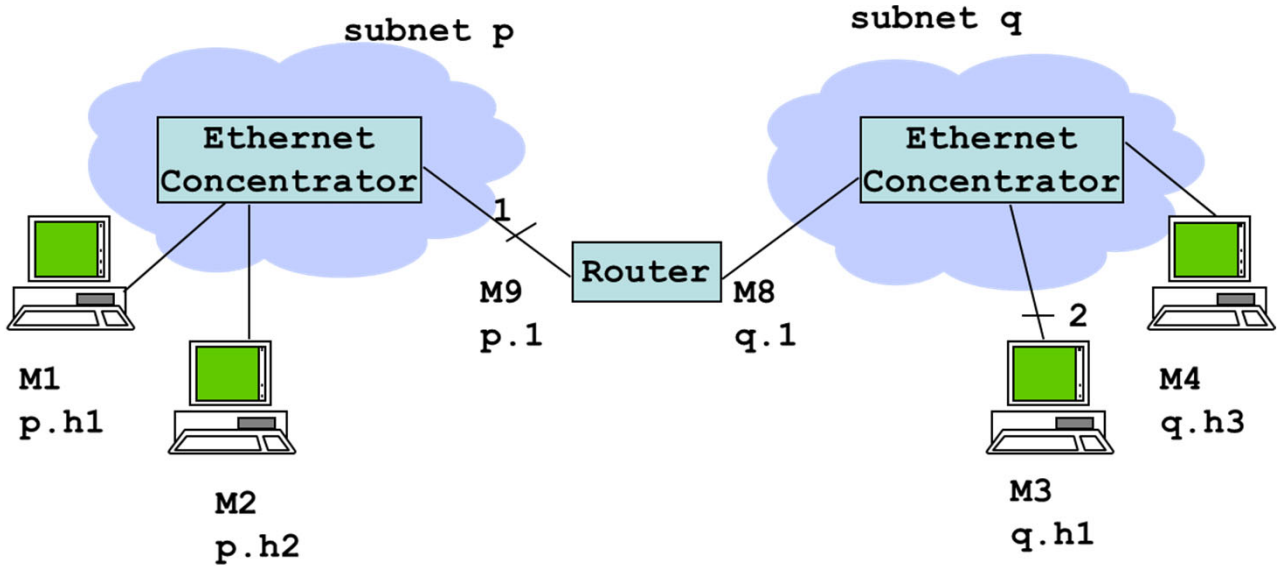
ARP request is broadcast to all nodes in LAN (instead of sent to solicited node multicast address)

Look inside an ARP packet

Ethernet II
- Destination: ff:ff:ff:ff:ff:ff (ff:ff:ff:ff:ff:ff)
- Source: 00:03:93:a3:83:3a (Apple_a3:83:3a)
- Type: ARP (0x0806)
- Trailer: 000000000000000000000000000000000...

Address Resolution Protocol (request)
- Hardware type: Ethernet (0x0001)
- Protocol type: IP (0x0800)
- Hardware size: 6
- Protocol size: 4
- Opcode: request (0x0001)
- Sender MAC address: 00:03:93:a3:83:3a (Apple_a3:83:3a)
- Sender IP address: 129.88.38.135 (129.88.38.135)
- Target MAC address: 00:00:00:00:00:00 (00:00:00_00:00:00)
- Target IP address: 129.88.38.254 (129.88.38.254)

M1 sends a packet to M3 for the first time since last reboot.



A. M1 sends an NS /ARP packet for q.h1
B. M1 sends an NS / ARP packet for p.1
C. None of the above
D. I don't know

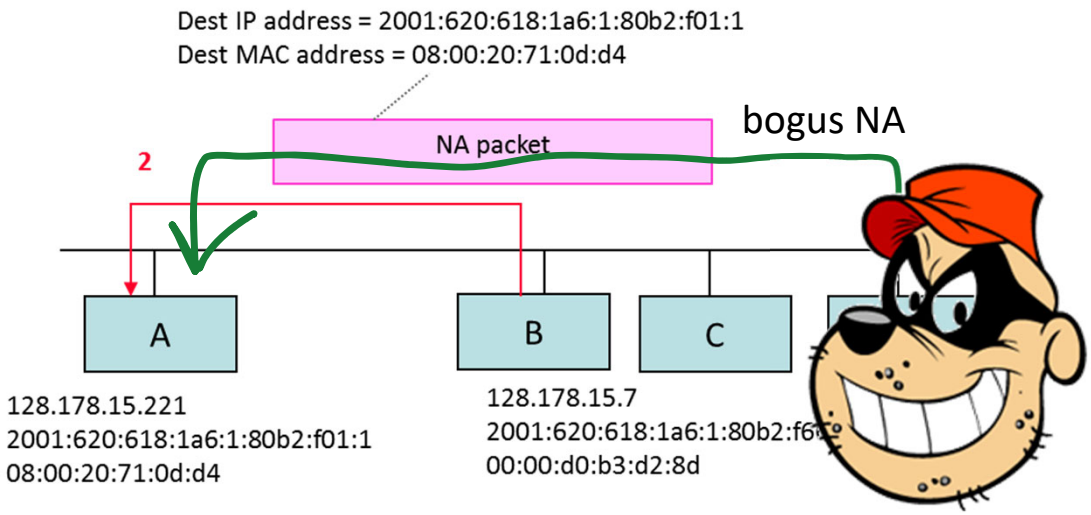
Solution

Answer B: Since M3 is not in the same subnet, M1 needs to find the MAC address of its default router, namely p.1.

Note that the IP address of the default router such as p.1. is in M1's configuration, but not the MAC address of the default router.

Security Issues with ARP/ NDP

ARP requests / replies may be falsified (ARP spoofing). Attacker will capture all packets intended to B (e.g. man in the middle attack)



DHCP Snooping and Dynamic ARP Inspection can prevent ARP spoofing in LANs

DHCP snooping = switch/Ethernet concentrator/WiFi base station observes all DHCP traffic and remembers mappings IP addr ↔ MAC addresses
(DHCP is used to automatically configure the IP address at system startup)

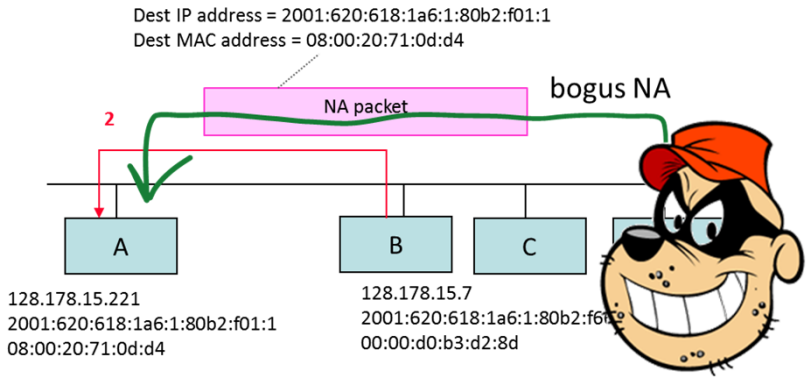
Dynamic ARP inspection: switch filters all ARP (or NDP) traffic and allows only valid answers – removes broadcasts (IPv4) and multicasts (IPv6)

Such solutions are deployed in enterprise networks, rarely in homes or WiFi access points

Secure NDP (SEND)

What ?

Make NDP spoofing impossible.



How ?

Host B has public/private key pair P/p.

EUI of B is a CGA (cryptographically generated address) = secure hash of P and IPv6 address prefix (and other fields such as counters). This binds EUI to P.

NA message sent in response to A contains a signature (RSA) computed with p. A can verify signature using P (which is public). Only the owner of p (which is secret) can send a valid NA.

Solves the problem but not widespread yet – requires a strong hash function (stronger than SHA1 as in current's version).

A private/public key system such as RSA or ECDSA has two keys : one public and one private (secret). With RSA, a clear text message can be encrypted with the private key and decrypted with the public key (or vice versa).

B signs the NA sent to A by using the private key p. A can verify the signature by applying RSA decryption with the public key P. This proves that the NA was originated by a system that knows the private key p.

A can also verify that B’s EUI is derived from the public key P, since the hash algorithm is known and public.

Anyone can generate B’s EUI but only the owner of p can send a valid NA.

6. Host Configuration

An IP host needs to be configured on each interface with

- IP address of this interface
- Mask of this interface
- IP address of default router
- IP address of DNS server

Can be done manually, or automatically with

- IPv4 → DHCP (Dynamic Host Configuration Protocol)
- IPv6 → DHCP stateful, SLAAC (stateless), DHCP stateless

Same applies to routers connected to a provider

- IPv4 → PPP
- IPv6 → PPP, DHCP with Prefix Delegation

DHCP with IPv4

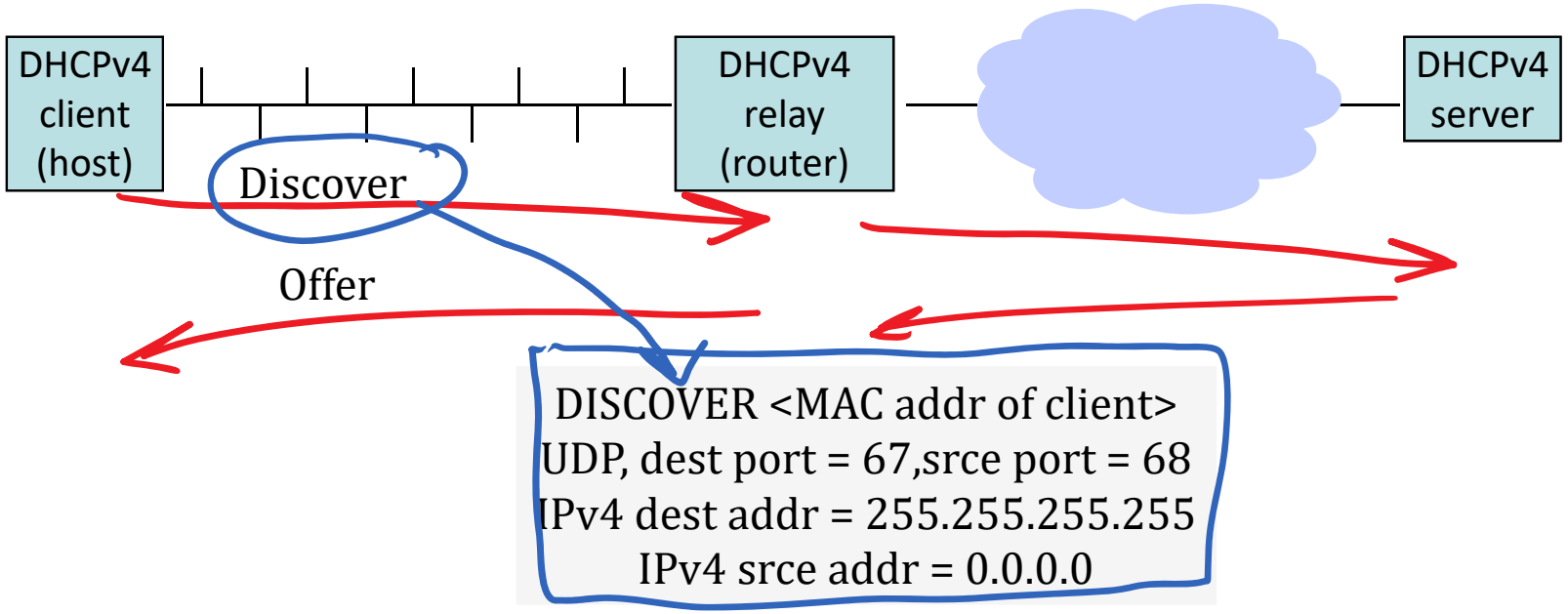
Configuration info is kept in central DHCP server, contacted by host when it needs an IP address; is commonly used with IPv4. Also works with IPv6 (with modifs – called stateful DHCP).

Problem: host cannot contact DHCP server since it is still does not have an IP address;

Solution: router implements a "DHCP Relay" function.

Two phase commit to avoid inconsistent reservations.

Limited lifetime - renewals



Stateless Address Autoconfiguration (SLAAC)

Why invented: avoid configuring DHCP servers - fully automatic

How it works :

1. host auto-configures a link local address; 64 bit host part obtained by one of these methods:

   manually assigned e.g. ::1;
   algorithmically derived from MAC address ("modified EUI 64")
   randomly assigned (RFC7217) to make it opaque
   cryptographically generated address (CGA) –hash public key of host

2. host performs address duplication test by sending a multicast packet (to solicited node multicast address)

3. host tries to also obtain globally valid addresses by obtaining network prefix from routers if any present; prefix needs not be 64 bits, a new host part is computed with same methods.

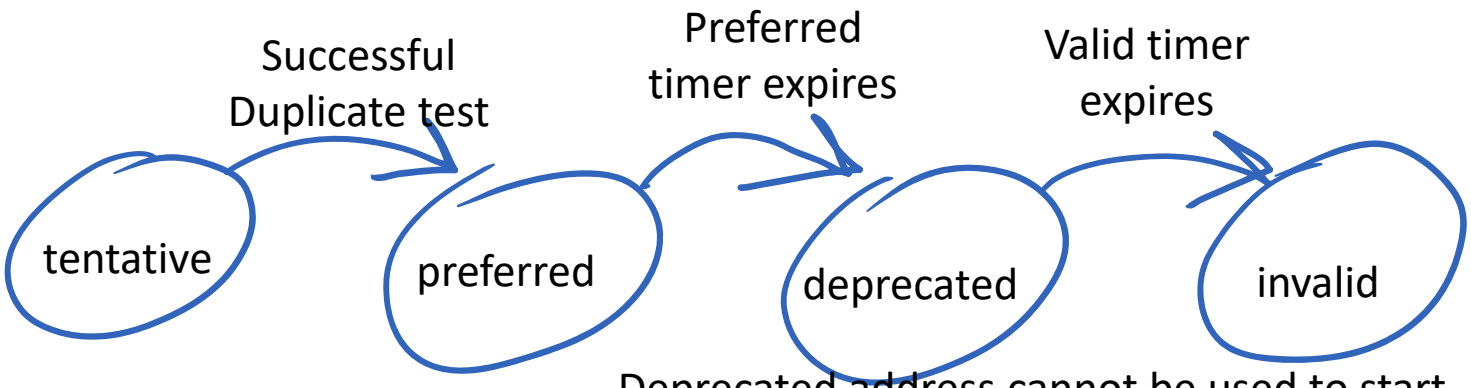
Temporary IPv6 Host Addresses

Problem: MAC address allows tracking

Randomly assigned Host Part can be used as alternative

Host randomly computes one tentative host part

- Duplicate test is used to avoid (unlikely) collisions
- Has a limited lifetime
- Limited lifetime, renewed before expiration



- Deprecated address cannot be used to start new TCP connections
- Host should obtain a new address

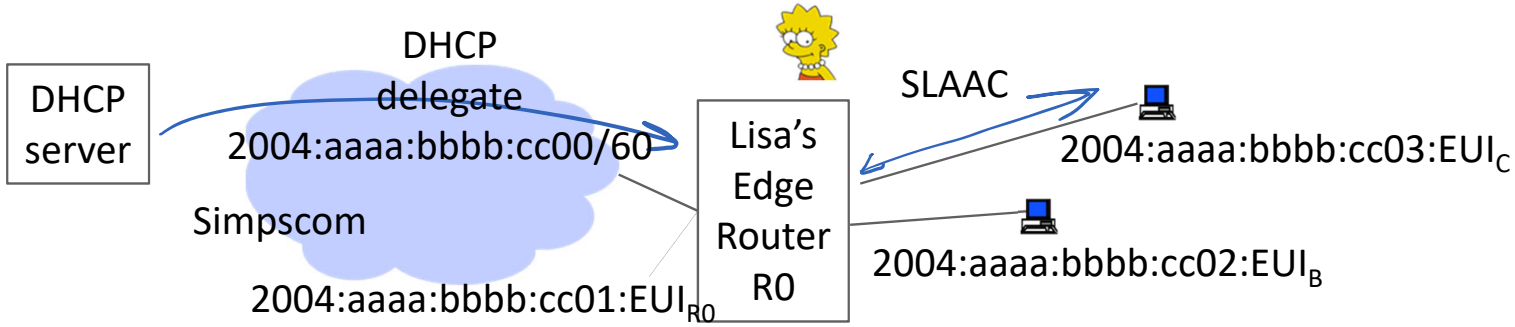
Other Bells and Whistles

The Point to Point Protocol (PPP): allocates address automatically over telecom lines (modem, ADSL); for v4 and v6

Stateless DHCP: gives DNS server address to host, used after SLAAC.

Router Advertisements: router on link indicates to host address of DNS server (RFC 6106), used after SLAAC

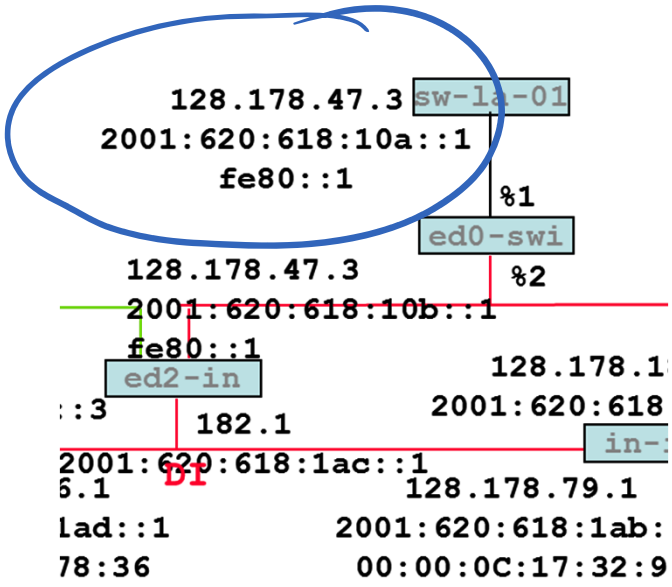
DHCP with Prefix Delegation



Why ? A home (or enterprise) IPv6 router R0 is configured by ISP using DHCP. Local devices are autoconfigured from home router using e.g. SLAAC. Home router needs an IPv6 prefix for the entire home network.

How ? ISP DHCP server (delegating router) provides to home router not just its IPv6 address but also the network prefix that this router can delegate to its devices. This is called prefix delegation. This prefix may include the prefix of the link from ISP to R0 (RFC 6603).

Multiple Addresses per Interface are the Rule with IPv6



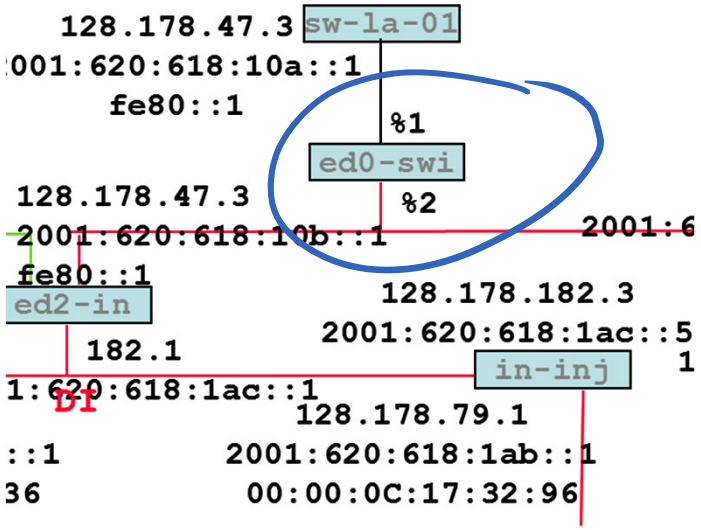
A host interface typically has

- One or several link local addresses
- Plus one or several global unicast addresses (secure (CGA) address, temporary addresses)

The preference selection algorithm, configured by OS, says which address should be used as source address – see RFC 3484

In contrast, there is usually only one *IPv4* address per interface

Zone Index



Identifies an interface inside one machine that has several interfaces – typically visible in Windows machines

Never inside an IP packet

E.g. fe80::1%2 means: the destination IPv6 address fe80::1 on interface %2

IP Configuration Example

Wireless LAN adapter Wireless Network Connection:

Physical Address. . . . . : 10-0B-A9-A3-91-08
DHCP Enabled. . . . . : Yes
Autoconfiguration Enabled . . . . : Yes
Link-local IPv6 Address . . . . . : fe80::945c:d22c:b0e2:a885%16(Preferred)
IPv4 Address. . . . . : 123.255.96.194(Preferred)
Subnet Mask . . . . . : 255.255.252.0
Lease Obtained. . . . . : mercredi 29 juillet 2020 09:05:03
Lease Expires . . . . . : mercredi 29 juillet 2020 09:35:02
Default Gateway . . . . . : 123.255.99.254
DHCP Server . . . . . : 10.3.1.12
DHCPv6 IAID . . . . . : 386927529
DHCPv6 Client DUID. . . . . : 00-01-00-01-16-E8-19-59-F0-DE-F1-BE-ED-EB
DNS Servers . . . . . : 202.45.188.37
137.189.192.3
137.189.196.3
NetBIOS over Tcpip. . . . . : Enabled

IAID = logical number of this interface, assigned by client

Ethernet MAC address
Identifies this host in the DHCP database

IPv4 Link Local Addresses

Some form of autoconfiguration also exists with IPv4

When host boots, if no DHCP and no configuration info available, it picks an IPv4 link local address at random in the 169.254/16 block

Address duplicate test is performed by broadcast

Allows to operate in routerless network («Dentist's Office», à la AppleTalk) but not in a general setting

Implemented in Windows, not supported by the Linux version we use in the lab

When an IPv4 host uses DHCP, which of the following information does it acquire:

A. its IP address;

B. its subnet mask

C. its default gateway address

D. its DNS server address

A. A
B. A, B
C. A, B, C
D. A, B, C, D
E. None of the above
F. I don’t know

With SLAAC an IPv6 host has…

A. A link local address and, if a router is present in the subnet, also a global unicast address
B. If a router is present in the subnet a global unicast address and no link-local address
C. None of the above
D. I don’t know

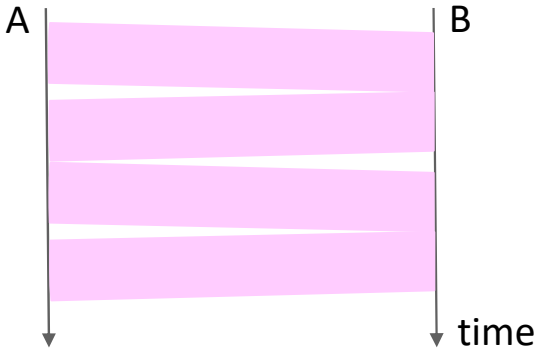
Solution

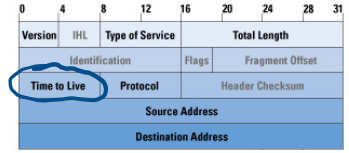
Answer D

Answer A

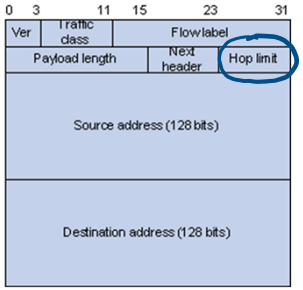
7. Hop Limit (HL) / Time to Live (TTL)

Why ? Avoid looping packets in transient loops. If propagation time is small compared to transmission time, a single packet caught in a loop can congest the line.





Transient loops may exist due to changes to routing tables + propagation latency



How ? Every IP packet has a field on 8 bits (from 0 to 255) (called Hop Limit for IPv6 / Time To Live IPv4) that is decremented at every hop. When it reaches 0, packet is discarded. At source, value is 64 in principle.

Traceroute

Sends a series of packets (using UDP) with TTL = 1, 2, 3, …

tracert (windows) similar but uses ICMP

Routers on the path discard packets and send ICMP error message back to source
Source learns address of router on the path by looking at source address of error message

```
Tracing route to www.google.com [2a00:1450:4008:800::1012]
over a maximum of 30 hops:

 1    1 ms    <1 ms    <1 ms  cv-ic-dit-v151-ro.epfl.ch [2001:620:618:197:1:80b2:9701:1]
 2   <1 ms    <1 ms    <1 ms  cv-gigado-v100.epfl.ch [2001:620:618:164:1:80b2:6412:1]
 3   <1 ms    <1 ms    <1 ms  c6-ext-v200.epfl.ch [2001:620:618:1c8:1:80b2:c801:1]
 4    1 ms    <1 ms    <1 ms  swiEL2-10GE-3-2.switch.ch [2001:620:0:ffdc::1]
 5   <1 ms    <1 ms    <1 ms  swiLS2-10GE-1-2.switch.ch [2001:620:0:c00c::2]
 6    7 ms     7 ms     7 ms  swiEZ1-10GE-2-7.switch.ch [2001:620:0:c03c::2]
 7    8 ms     8 ms     7 ms  swiEZ2-P2.switch.ch [2001:620:0:c0c3::2]
 8    8 ms     8 ms     8 ms  swiIX2-P1.switch.ch [2001:620:0:c00a::2]
 9    8 ms     8 ms     8 ms  swissix.google.com [2001:7f8:24::4a]
10   38 ms    34 ms    15 ms  2001:4860::1:0:4ca2
11   14 ms    14 ms    17 ms  2001:4860::8:0:5038
12   17 ms    50 ms    17 ms  2001:4860::8:0:8f8e
13   24 ms    24 ms    24 ms  2001:4860::8:0:6400
14   25 ms    25 ms    25 ms  2001:4860::1:0:6e0f
15   25 ms    24 ms    25 ms  2001:4860:0:1::4b
16   25 ms    25 ms    25 ms  ber01s08-in-x12.1e100.net [2a00:1450:4008:800::1012]
```

Other fields in IP Header

Type of service / Traffic Class

- Differentiated Services (6bits) – sort of priority eg voice over IP
  Used only in corporate networks
- Explicit Congestion Notification (2bits) see congestion control

Total length / Payload length

- in bytes including header
- ≤ 64 Kbytes; limited in practice by link-level MTU (Maximum Transmission Unit)
- every subnet should forward packets of 576 = 512 + 64 bytes

Protocol / Next Header = identifier of protocol

- 6 = TCP, 17 = UDP
- 1 = ICMP for IPv4, 58 = ICMP for IPv6
- 4 = IPv4; 41 = IPv6 (encapsulation = tunnels)
- 50 = ESP (encrypted payload)
  51 = AH (authentication header)

Checksum

- IPv4 only, protects header against bit errors
- Absent in IPv6 ⇒ layer 2 and router hardware assumed to have efficient error detection

ICMP is used to carry error messages

A host generates a packet with Hop Limit = 1

A. This packet is invalid
B. This packet will never be forwarded by a bridge nor by a router
C. This packet will never be forwarded by a bridge but may be forwarded by a router
D. This packet will never be forwarded by a router but may be forwarded by a bridge
E. None of the above is true
F. I don’t know

Solution

Answer D

This packet cannot be forwarded by a router because it would decrement the HL and obtain 0. It can be forwarded by a bridge because a bridge does not examine the IP header.

Note that such a packet is perfectly valid. Sources put HL=1 when they want to be sure that the packet remains in the LAN.

Conclusion

IP is built on two principles:

- one IP address per interface and longest prefix match; this allows to compress routing tables by aggregation
- inside subnet, don’t use routers

IPv4 and IPv6 are not compatible – interworking requires tricks

NATs came as an after-thought and are widely deployed

ARP/NDP finds the MAC address corresponding to an IP address

DHCP is used allocates IP address, network mask and DNS server’s IP address to a host; SLAAC automatically allocates IPv6 addresses without DHCP

TTL/HL limits the number of hops of an IP packet